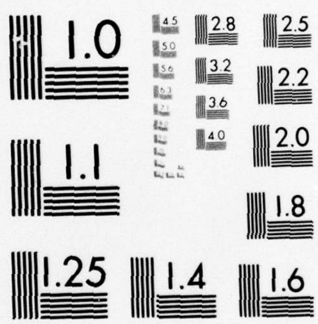
AD-A039 041 MOORE SCHOOL OF ELECTRICAL ENGINEERING PHILADELPHIA P--ETC F/G 9/5
COEFFICIENT AND DATA QUANTIZATION IN MATCHED FILTERS FOR DETECT--ETC(U)
1977 T L LIM, S A KASSAM AF-AFOSR-3154-77
UNCLASSIFIED AFOSR-TR-77-0555 NL

| OF |

AD
A039041

END
DATE
FILMED
5-77



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

COEFFICIENT AND DATA QUANTIZATION IN MATCHED FILTERS FOR DETECTION

Tong Leong Lim and Saleem A. Kassam
The Moore School of Electrical Engineering
University of Pennsylvania
Philadelphia, PA. 19104

Abstract

Quantization of data and representation of the coefficients is studied in digital matched filters for weak-signal detection. An algorithm for optimum coefficients and equations for the optimum input quantizer are obtained for the known signal in additive noise problem. Some numerical performance results are given.

I. INTRODUCTION

A matched filter is often used as a detector for testing a hypothesis about an input signal; it is the optimal processor in this role for Gaussian input noise, and may also be considered as the optimum processing scheme in non-Gaussian noise in the weak-signal case. We will elaborate on this in the next section; the general theory and applications of matched filters may be found, for example, in [1,2].

In this paper we will be concerned with digital matched filters operating on discrete-time data, and we will examine the effects of, and optimization with respect to, finite-bit representation or quantization of the coefficients and analog inputs. Most previous investigations in this direction have been concerned either with data quantization only (e.g. [3]) or have assumed only simple one-bit coefficient representations [2].

In the next section we briefly consider the basic results on local, or weak-signal, detection of signals based on quantized data and finite-bit coefficients. In Section III implementation of optimum digital matched filtering is considered, and an algorithm for determining the best coefficient representation is discussed. In Section IV we give some numerical examples.

II. PERFORMANCE WITH QUANTIZED COEFFICIENTS AND INPUTS

Let us assume that an observation vector $\underline{X} = (X_1, X_2, \ldots, X_n)$ is available, and is described by the equation

$$X_i = \theta s_i + N_i, \quad i=1,2,\ldots,n, \quad \theta \geq 0. \quad (1)$$

Here the vector $\underline{s} = (s_1, s_2, \ldots, s_n)$ is a known signal vector, θ is the amplitude of the signal, and the vector $\underline{N} = (N_1, N_2, \ldots, N_n)$ is a vector of independent identically distributed noise samples each with symmetric density f.

If we consider as a detection statistic for testing $H_0: \theta=0$ vs. $H_1: \theta>0$ a

$$T = \sum_{i=1}^{n} g_i(X_i), \quad (2)$$

and use as a criterion of optimality the differential signal-to-noise ratio

$$\text{DSNR} = \frac{\left[\frac{d}{d\theta} E\{T\}\big|_{\theta=0}\right]^2}{\text{Var}\{T\}_{\theta=0}} \quad (3)$$

we find (from the Schwarz inequality) that the optimum T maximizing DSNR is

$$T_{opt} = \sum_{i=1}^{n} s_i \, g_{opt}(X_i) \quad (4)$$

where

$$g_{opt}(X_i) = -f'(X_i)/f(X_i) \quad (5)$$

the prime denoting the first derivative of the function. The criterion of (3) is a reasonable one for the weak-signal case, being a modified case of the usual SNR criterion [2]. It is also well-known, of course, that T_{opt} is the locally-optimum statistic maximizing the slope of the power function for testing H_0 vs. H_1 [4].

In the case of Gaussian noise T_{opt} is the output of a linear filter matched to the signal vector $\underline{s}$. In the general case, we may consider T_{opt} to be the output of a similar filter preceded by the instantaneous nonlinearity g_{opt}. For the digital matched filter both the coefficients, the s_i, and the data inputs, the $g_{opt}(X_i)$, have to be replaced by suitably quantized versions r_i and $q(X_i)$, respectively.

Presented at the Conference on Information Sciences and Systems, The Johns Hopkins University, March 1977; to be published in the Proceedings of the Conference.

9 (4) See 1473 in back

To optimize the performance of the digital matched filter, then, we have to optimize DSNR of (3) for the case where the detection statistic is

$$T_d = \sum_{i=1}^{n} r_i\, q(X_i), \tag{6}$$

the optimization being with respect to a choice of the r_i and the input quantizer q, given $2^b=2k$ levels or b bits for coefficient representation and $2^t=2m$ levels or t bits for the data quantization. Note that (6) represents a general digital matched filter operating on quantized data, for any kind of input noise density.

Applying (3) to T_d, we find that DSNR for T_d with an odd-symmetric, even-state quantizer q is given by

$$DSNR_d = \frac{\left[\sum_{i=1}^{n} s_i r_i\right]^2}{\sum_{i=1}^{n} r_i^2} \cdot \frac{2\left[\sum_{j=1}^{m} \ell_j [f(a_j)-f(a_{j-1})]\right]^2}{\sum_{j=1}^{m} \ell_j^2 [F(a_{j-1})-F(a_j)]} \tag{7}$$

where F is the distribution function corresponding to f. In (7) the ℓ_j are the output levels for positive x of the symmetric quantizer q, with q(x) being ℓ_j whenever x is in the interval (a_j, a_{j-1}). The breakpoints a_j satisfy $a_j < a_{j-1}$, and by definition $a_m=0$ and $a_0=\infty$.

The problem is to maximize the quantity

$$J \triangleq \frac{\left[\sum_{i=1}^{n} s_i r_i\right]^2}{\sum_{i=1}^{n} r_i^2}, \tag{8}$$

which is a factor in $DSNR_d$. In addition, for a given set of levels ℓ_j for the quantizer q, the best set of breakpoints a_j may be obtained. Then the overall performance of the digital matched filter may be optimized with respect to allocation of b+t bits between coefficient and data quantization.

III. PERFORMANCE OPTIMIZATION

As discussed in the previous section, the objective is to maximize $DSNR_d$ of (7). The two factors in (7) are decoupled, so that the coefficient representation and data quantization problems can be treated independently of each other. We first consider maximization of the coefficient factor J defined in (8). This quantity can be written as

$$J = \frac{(\underline{r}\cdot\underline{s})^2}{||\underline{r}||^2} \tag{9}$$

where $\underline{s}$ is the previously defined signal vector and $\underline{r} = (r_1, r_2, \ldots, r_n)$, the vector of filter coefficients by which the reference signal $\underline{s}$ is represented. If we let ϕ be the angle between $\underline{s}$ and $\underline{r}$, so that

$$\cos^2\phi = \frac{(\underline{r}\cdot\underline{s})^2}{||\underline{r}||^2\,||\underline{s}||^2}, \tag{10}$$

we can express J as

$$J = ||\underline{s}||^2 \cos^2\phi . \tag{11}$$

Thus, maximization of J for a given reference signal is simply maximization of $\cos^2\phi$, and we therefore have to pick the vector $\underline{r}$ closest to $\underline{s}$.

Since it was not easy to get an analytic solution for the coefficient vector $\underline{r}$ maximizing $\cos^2\phi$ for a given $\underline{s}$ (and with a constraint on the number of bits, b, for coefficient representation), an efficient computer technique was developed which is described below.

It was assumed that the coefficient representation $\underline{r}$ is obtained through a quantizer with one of three possible ranges of $2^b=2k$ levels:

(a) The levels $0,1,\ldots,2k-1$

(b) The levels $-k+1, -k+2,\ldots,k$

(c) The levels $-k, -k+1,\ldots,k-1$

Range (a) is obviously to be used if $\underline{s}$ has positive components. A signal vector with all negative components may be complemented, and therefore be represented in this range also. Ranges (b) and (c) are more natural choices if positive as well as negative components are present in $\underline{s}$. Note that we are considering the class of quantizers with an even number of levels, one of which is the 0 level.

The search algorithm for $\underline{r}$ is described by the following sequence of steps. It is assumed that the signal vector components have been ordered, that is $s_1 \le s_2 \le \ldots \le s_n$, and are non-zero.

0. Initialization: $j = 2k-1$, $C=0$.
1. Check if $s_n>0$. If not, invert the sequence of components in $\underline{s}$ so that now $\underline{s} = (-s_n, -s_{n-1}, \ldots, -s_1)$, and go to step 3.
2. Check if $\sum_{i=1}^{n} s_i \ge 0$. If the sum is negative, invert the sequence of components in $\underline{s}$, so that now $\underline{s} = (-s_n, -s_{n-1}, \ldots, -s_1)$. Note that inverting the sequence in step 1 results in a sequence which fulfills the condition in this step.
3. Form the vector $\underline{s}(j) = (s_1(j), s_2(j), \ldots, s_n(j))$ where $s_i(j) = js_i/s_n$. Let $I_i(j)$ be the largest integer *less than, or equal to,* $s_i(j)$. *Thus* $I_n(j) = j$; the algorithm now looks for the first $(n-1)$ components of $\underline{r}(j) = (r_1(j), r_2(j), \ldots, r_{n-1}(j), j)$ so that the cosine squared of the angle between $\underline{s}$ and $\underline{r}(j)$ is maximized.
4. If $j=k$, go to step 6. If $j<k$, go to step 7.
5. If $s_i(j)>0$, consider the two possibilities $I_i(j)$ and $U_i(j)$ for $r_i(j)$, where $U_i(j)$ is the smallest integer larger than, or equal to, $s_i(j)$. If $s_i(j)<0$, set $r_i(j)=0$. Go to step 8.
6. If $s_i(j) \ge -k+1$, consider the possibilities $I_i(j)$ and $U_i(j)$ for $r_i(j)$. If $s_i(j) < k+1$, set $r_i(j) = -k+1$. Go to step 8.
7. If $s_i(j) \ge -k$, consider the possibilities $I_i(j)$ and $U_i(j)$ for $r_i(j)$. If $s_i(j) < -k$, set $r_i(j) = -k$.
8. Compute the square of the cosine, $c^2(j)$, of the angle between $\underline{s}$ and $\underline{r}(j)$ for each different possible combination of components of $\underline{r}(j)$; let $\underline{r}_m(j)$ be the vector yielding a maximum value $c_m^2(j)$ for $c^2(j)$. If $c_m^2 > C$, assign the value $c_m^2(j)$ to C and define $\underline{r}$ to be $\underline{r}_m(j)$.
9. If $j>1$, set $j=j-1$ and go to step 3.
10. Stop. The optimum coefficient vector is $\underline{r}$, and C is the square of the cosine of the angle between $\underline{r}$ and $\underline{s}$, if the two conditions of steps 1 and 2 had been fulfilled. Otherwise the optimum coefficient vector is in inverse order in $\underline{r}$.

On an intuitive basis, one might choose the best coefficient vector in the following way for the simple case where the s_i are positive. Pick the vector of coefficients $\tilde{\underline{r}}$, where the components $\tilde{r}_i$ are the integers closest to the $s_i(2k-1)$. However, it is clear that $\tilde{\underline{r}}$ need not be an optimum set maximizing $\cos^2\phi$. This can be seen by considering the case b=1, giving two-level representation, for $(n-1)$ identical, small s_i and a large value for s_n. In the next section this is further illustrated for an example which is not as extreme.

We also need to consider optimization with respect to the data quantizer. For the symmetric even-state input quantizer we are considering, specification of the number of levels $2^t=2m$ (with t bits) fixes the levels at values $\pm 1, \pm 2, \ldots, \pm m$. With $\ell_j = m-j+1, j=1,2,\ldots,m$, the second factor K in (9), defined by

$$K = \frac{2\left[\sum_{j=1}^{m} \ell_j [f(a_j) - f(a_{j-1})]\right]^2}{\sum_{j=1}^{m} \ell_j^2 [F(a_{j-1}) - F(a_j)]}, \qquad (12)$$

can be maximized with respect to the a_j. The optimum set $\{a_j\}_{j=1}^{m-1}$ can easily be obtained for specific densities f. Setting the partial derivative of K with respect to a_j equal to zero, we find a necessary condition for the maximizing values of a_j:

$$\frac{-f'(a_j)}{f(a_j)} = \left[\frac{\ell_j + \ell_{j+1}}{2}\right] \frac{\sum_{j=1}^{m} \ell_j [f(a_j) - f(a_{j-1})]}{\sum_{j=1}^{m} \ell_j^2 [F(a_{j-1}) - F(a_j)]} \qquad j=1,\ldots,m-1 \tag{13}$$

Equation (13) may be solved for specific densities. For the case of a Gaussian density with variance σ^2, the equations reduce to

$$a_j = \sigma^2 \left[\frac{\ell_j + \ell_{j+1}}{2}\right] L, \; j=1,\ldots,m \tag{14}$$

where

$$L = \frac{\sum_{j=1}^{m} \ell_j [f(a_j) - f(a_{j-1})]}{\sum_{j=1}^{m} \ell_j^2 [F(a_{j-1}) - F(a_j)]} . \tag{15}$$

Substitution of (14) in (15) leads to a single equation for L, which may then be solved (numerically) and hence the optimum a_j can be obtained. The next section contains some specific numerical results.

IV. RESULTS AND DISCUSSION

The considerations of the previous section were applied to several specific cases, two of which are presented here. A signal vector $\underline{s}$ of 10 components and one of 15 components is shown in Table I, together with optimum representations for one, two and three bits. The maximum values J_{max} of J [Equation (11)] normalized by $||\underline{s}||^2$ are also shown in Table I. The results indicate that even with only two bits for coefficient representation $(\cos^2\phi)_{max}$ is very close to unity, which corresponds to the analog case. The exact numerical values depend on the coefficient vector and its length.

In the second part of Section III the optimization of the second factor K in (7) was discussed. The optimum performance for the specific case of Gaussian noise with unit variance is shown in Table II. Again, in this case the value K=1 is obtained for unquantized data. Table III combines the results of Tables I and II, giving the overall performance for the different allocations of a fixed number of bits (six bits) between coefficient representation and data quantization. In general, the optimum allocation will depend on the length of the vector $\underline{s}$, its components and the type of noise density. It is seen that one need consider only a small number of bits to achieve near-analog performance.

In Section III an approximation method of quantizing the coefficient sequence $\underline{s}$ was also discussed. Table IV gives an example of this approximation method using the same sequence as in Table 1(a) with n=10. Note that the discrepancy is larger for the lower-order quantizers, as expected. In fact, the result of the approximation for 3 bits is identical to that in Table 1. It is seen that in this example, the approximation methods gives performance very close to optimum.

ACKNOWLEDGEMENT

This research is supported by the Air Force Office of Scientific Research, Air Force Systems Command, USAF, under Grant No. AFOSR-77-3154.

REFERENCES

1. H. L. Van Trees, Detection, Estimation and Modulation Theory, Part I, New York: Wiley, 1968.
2. G. L. Turin, "An introduction to digital matched filters," Proc. IEEE, vol. 64, pp. 1092-1112, July 1976.
3. S. A. Kassam, "Optimum quantization for signal detection," IEEE Trans. Commun., vol. COM-25, (to be published).
4. J. Capon, "On the asymptotic efficiency of locally-optimum detectors," IRE Trans. Inform. Theory, vol. IT-7, pp. 67-71, April 1961.

Table I: Optimum Coefficient Representation

(a) n=10

COEFFICIENTS	-1.82 1.58	-0.92 1.98	-0.08 2.30	0.77 2.41	1.22 3.10	$\frac{J_{max}}{\|\|\underline{s}\|\|^2} = (\cos^2\phi)_{max}$
ONE BIT REPRESENTATION	0 1	0 1	0 1	0 1	1 1	0.792
TWO BIT REPRESENTATION	-1 1	-1 2	0 2	1 2	1 2	0.957
THREE BIT REPRESENTATION	-2 2	-1 3	0 3	1 3	2 4	0.990

(b) n=15

COEFFICIENTS	-4.3 -0.7 2.1	-2.1 0.1 3.1	-1.8 0.3 3.3	-1.7 1.1 4.1	-1.2 1.6 4.5	$\frac{J_{max}}{\|\|\underline{s}\|\|^2} = (\cos^2\phi)_{max}$
ONE BIT REPRESENTATION	0 0 1	0 0 1	0 0 1	0 0 1	0 0 1	0.604
TWO BIT REPRESENTATION	-1 0 1	-1 0 2	-1 0 2	-1 1 2	-1 1 2	0.907
THREE BIT REPRESENTATION	-3 0 1	-1 0 2	-1 0 2	-1 1 3	-1 1 3	0.979

Table II: Optimum Data Quantization (Gaussian Noise, Unit Variance)

QUANTIZER BITS	1	2	3	4
OPTIMUM VALUE OF K	0.637	0.842	0.946	0.984

Table III: Bit Allocation Between Coefficient and Data Quantization (Gaussian Noise, Unit Variance)

(a) n=10

Coefficient Bits	1	2	3	4	5
Data Bits	5	4	3	2	1
$\frac{DSNR}{\|\|\underline{s}\|\|^2}$	<0.792	0.941	0.936	<0.842	<0.637

(b) n=15

Coefficient Bits	1	2	3	4	5
Data Bits	5	4	3	2	1
$\frac{DSNR}{\|\|\underline{s}\|\|^2}$	<0.604	0.892	0.926	<0.842	<0.637

Table IV: Suboptimum Coefficient Representation

COEFFICIENTS	-1.82 1.58	-0.92 1.98	-0.08 2.30	0.77 2.41	1.22 3.10	$\frac{J_{max}}{\|\|\underline{s}\|\|^2} = (\cos^2\phi)_{max}$
ONE BIT REPRESENTATION	0 1	0 1	0 1	0 1	0 1	0.775
TWO BIT REPRESENTATION	-1 1	-1 1	0 1	0 2	1 2	0.930
THREE BIT REPRESENTATION	-2 2	-1 3	0 3	1 3	2 4	0.990

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (*When Data Entered*)

REPORT DOCUMENTATION PAGE		READ INSTR[UCTIONS] BEFORE COMPLETING [FO]RM
1. REPORT NUMBER AFOSR - TR- 77 - 0555	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (*and Subtitle*) COEFFICIENT AND DATA QUANTIZATION IN MATCHED FILTERS FOR DETECTION		5. TYPE OF REPORT & PERIOD COVERED Interim 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Tong Leong Lim Saleem A. Kassam		8. CONTRACT OR GRANT NUMBER(s) AFOSR 77-3154
9. PERFORMING ORGANIZATION NAME AND ADDRESS University of Pennsylvania Dept of Systems Engineering Philadelphia, PA 19104		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 2304/A5
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research/NM Bolling AFB, Washington, DC 20332		12. REPORT DATE 1977 13. NUMBER OF PAGES 6
14. MONITORING AGENCY NAME & ADDRESS(*if different from Controlling Office*)		15. SECURITY CLASS. (*of this report*) UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (*of this Report*)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (*of the abstract entered in Block 20, if different from Report*)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (*Continue on reverse side if necessary and identify by block number*)

20. ABSTRACT (*Continue on reverse side if necessary and identify by block number*)

Quantization of data and representation of the coefficients is studied in digital matched filters for weak-signal detection. An algorithm for optimum coefficients and equations for the optimum input quantizer are obtained for the known signal in additive noise problem. Some numerical performance results are given.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: 18 AFOSR - TR - 77 - 0555 19	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): 6 COEFFICIENT AND DATA QUANTIZATION IN MATCHED FILTERS FOR DETECTION	5. TYPE OF REPORT & PERIOD COVERED: 9 Interim rept.
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): 10 Tong Leong Lim, Saleem A. Kassam	8. CONTRACT OR GRANT NUMBER(s): 15 AF- AFOSR-77-3154 - 77
9. PERFORMING ORGANIZATION NAME AND ADDRESS: University of Pennsylvania, Dept of Systems Engineering, Philadelphia, PA 19104	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 16 61102F 2304/A5 17
11. CONTROLLING OFFICE NAME AND ADDRESS: Air Force Office of Scientific Research/NM, Bolling AFB, Washington, DC 20332	12. REPORT DATE: 11 1977
	13. NUMBER OF PAGES: 6 12 8p.
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report): UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

409 814

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Quantization of data and representation of the coefficients is studied in digital matched filters for weak-signal detection. An algorithm for optimum coefficients and equations for the optimum input quantizer are obtained for the known signal in additive noise problem. Some numerical performance results are given.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

6pp